МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В. ЛОМОНОСОВА

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Котляров Олег Леонидович

Методы экстраполяции нерегулярных временных рядов

Специальность 01.04.02 – «Теоретическая физика»

Автореферат

диссертация на соискание ученой степени кандидата
физико-математических наук

Москва – 2006

Работа выполнена на физическом факультете Московского государственного университета им. М.В. Ломоносова

Научный руководитель
доктор физ.-мат. наук, профессор А.Ю. Лоскутов

Официальные оппоненты:
доктор физ.-мат. наук, профессор Г.Э. Норман
доктор физ.-мат. наук, профессор А.И. Чуличков

Ведущая организация
Российская экономическая академия им. Г.В. Плеханова

Защита состоится «___» ________ 2006 года в _____ часов на заседании диссертационного совета К 501.001.17 физического факультета МГУ им. М.В. Ломоносова по адресу 119992, Москва, МГУ, физический факультет, ауд. _________.

С диссертацией можно ознакомиться в библиотеке физического факультета МГУ им. М.В. Ломоносова.

Автореферат разослан «___» ________ 2006 года.

Ученый секретарь
диссертационного совета
доктор физ.-мат. наук П.А. Поляков

Общая характеристика работы

Актуальность темы

Оценка результатов любого эксперимента базируется на обработке полученных данных. В этих условиях обработка временных рядов с целью извлечения из них полезной информации становится одной из важнейших задач любого исследования. Но этим цели обработки временных рядов не ограничиваются. Достаточно часто определяющим является не только изучение свойств системы, породившей временной ряд, но – иногда и в первую очередь – прогноз дальнейшей динамики временного ряда, т.е. его экстраполяция. В метеорологии практический интерес представляет прогноз погоды на ближайшее время. В геофизике задача, имеющая на сегодня наибольшую общественную значимость, – это предсказание землетрясений. Подобное задачи прогнозирования возникают при изучении солнечной активности в астрофизике, в финансовом анализе при прогнозе курсов акций, биржевых индексов, а также во многих других областях исследований и научных дисциплинах.

С другой стороны, стремление заглянуть в будущее возникло у человечества очень давно, а вместе с ним возникли и различные способы постижения будущего. Однако область приложения методов и моделей, рассматриваемых в работе, далека от сфер приложения астрологии и хиромантии, а цели применения этих методов и моделей не столь глобальны. Основная задача, на которую они ориентированы, – исходя из имеющихся данных в предположении неизменности характера динамики системы, построить прогноз поведения наблюдаемой.

В основе большинства методов, связанных с обработкой временных рядов, лежит использование многомерного представления временного ряда в виде матрицы задержек – набора копий временного ряда, взятых с определенными лагами. Новым результатом теории динамических систем явилось установление факта, что пространство задержек при соблюдении определенных условий может рассматриваться как реконструкция фазового пространства нелинейной динамической системы, породившей временной ряд. Таким образом, была доказана возможность описания динамики многомерной системы по временному ряду наблюдаемой. В свою очередь, возможность описания и реконструкции динамики системы при определенных условиях позволяет прогнозировать ее дальнейшее поведение.

В рамках теории динамических систем было разработано достаточно много методов анализа и прогнозирования временных рядов. В представленной работе подробно рассмотрены два из них, отражающие два основных подхода к описанию динамики временных рядов: глобальный – метод сингулярного спектрального анализа – и локальный – метод локальной аппроксимации. Метод сингулярного спектрального анализа позволяет сгладить исходный ряд, снизить уровень случайных возмущений, выявлять периодические составляющие ряда и во многих случаях прогнозировать дальнейшее изменение изучаемой временной зависимости. Преимущества локальной аппроксимации проявляются в первую очередь при прогнозировании нерегулярных (хаотических и квазипериодических) стационарных временных рядов.

На сегодняшний день существует несколько вариантов метода локальной аппроксимации, которые различаются конкретным видом используемых моделей, способом прогнозирования на несколько шагов вперед и особенностями численных расчетов. Также имеется несколько модификаций сингулярного спектрального анализа. Однако, в отличие от вариантов метода локальной аппроксимации, они различаются не столь принципиально и достаточно подробно рассмотрены в литературе.

Таким образом, в первую очередь возникает задача систематизации существующих разновидностей метода локальной аппроксимации и выбора из них наиболее оптимальной в каждом конкретном случае. Кроме того, при исследовании реальных систем, как правило, приходится иметь дело с зашумленными данными, что может затруднять использование метода локальной аппроксимации. Решение перечисленных задач, как представляется, должно способствовать расширению сферы применения рассматриваемых методов экстраполяции временных рядов.

Цели диссертационной работы

- Разработка математической модели метода локальной аппроксимации, позволяющей выбирать оптимальный вариант метода исходя из характеристик исследуемого временного ряда.
- Анализ особенностей применения различных вариантов методов локальной аппроксимации и сингулярного спектрального анализа.
- Исследование возможностей применения методов локальной аппроксимации и сингулярного спектрального анализа при обработке временных рядов с аддитивным шумом.
- Разработка алгоритма выбора параметров локальной аппроксимации, не требующего визуального контроля и принятия решения исследователем.

Научная новизна

1. Исследован, расширен и систематизирован набор вариантов метода локальной аппроксимации.
2. Оценено качество прогнозов, получаемых различными вариантами метода локальной аппроксимации.
3. Предложен способ выбора оптимального варианта метода локальной аппроксимации.
4. Разработан метод снижения влияния шума на результаты прогноза.
5. Предложен критерий выбора параметров локальной аппроксимации.

Научная и практическая ценность

1. Предложен математический аппарат для аналитического исследования результатов прогноза.
2. Разработан программный комплекс, реализующий методы локальной аппроксимации и сингулярного спектрального анализа, в том числе с автоматическим выбором параметров и контролем получаемых результатов.
3. Выработаны рекомендации по численной реализации методов локальной аппроксимации и сингулярного спектрального анализа для обработки временных рядов.

Защищаемые положения

1. Общая математическая модель метода локальной аппроксимации и ее следствия.
2. Принцип предварительной SSA-фильтрации временного ряда для экстраполяции ряда по методу локальной аппроксимации.
3. Способ автоматического выбора параметров локальной аппроксимации.

Публикации

Основные результаты диссертационной работы изложены в восьми публикациях, в том числе пяти рецензируемых, перечень которых приведен в заключительной части автореферата, и доложены на ряде научных семинаров, а также на XI Всероссийской научной школе «Нелинейные волны – 2002», Нижний Новгород, «The 2nd Shanghai International Symposium on Nonlinear Science and Applications», 2005, Shanghai, China и «2nd International Nonlinear Science Conference», 2006, Heraklion, Crete, Greece.

Содержание работы

Диссертационная работа состоит из введения, 4 глав и заключения, включает 21 рисунок, список литературы из 114 наименований и одно приложение.

Во ***Введении*** обосновывается актуальность темы работы, сформулированы цели работы, кратко охарактеризованы научная новизна и практическая ценность полученных результатов и разработанных методик. Дана краткая аннотация каждой главы диссертационной работы.

Первая глава представляет собой основную часть литературного обзора работы. Здесь, в частности, проводится сравнение двух крайних подходов к описанию нерегулярности во временных рядах – описания в предположении нелинейного детерминированного поведения и линейного стохастического поведения. Эти два подхода рассматриваются наиболее часто из-за их ясного математического представления и наибольшего понимания и определяют границы проявления свойств нелинейности и стохастичности. На практике эти подходы оказываются взаимодополняющими: реальный мир временных рядов содержит все виды смесей нелинейности и стохастичности. Многие реальные источники нерегулярных сигналов, например, земная атмосфера, как установлено, являются принципиально нелинейными, но в некоторых случаях их эволюция может в результате усреднения сводиться к величинам, хорошо аппроксимируемым моделями типа ARMA. Для выявления «нелинейности» могут применяться специальные статистические тесты, которые заключаются, как правило, в формулировке некоторой нулевой гипотезы для порождающего процесса, которая может быть принята либо отклонена исходя из оценок выбранного нелинейного параметра.

Первая глава включает в себя также краткое изложение базовых принципов, подходов и определений теории динамических систем, в рамках которой рассматриваются теоретические основы обработки временных рядов.

Почти все методы анализа временных рядов, стандартные линейные или нелинейные, работают только для стационарных рядов. Это связано с тем, что любые изменения в динамике процесса во время измерений обычно сильно затрудняют анализ результатов. Тем самым вводятся определенные ограничения на обрабатываемые временные ряды. Для проверки стационарности ряд могут применяться специальные тесты на стационарность (нестационарность).

Использование теоретически обоснованных подходов при обработке реальных временных рядов сталкивается с весьма серьезной проблемой. Дело в том, что все теоретические результаты получены в предположении достижимости сколь угодно малых масштабов длины. А это возможно только при бесконечном количестве доступных данных. Кроме того, предполагается, что наблюдения доступны с произвольной точностью. Поэтому требуются специальные исследования применимости полученных теоретических результатов при наличии шума и в условиях конечной длины ряда.

Та же проблема возникает и при оценке количественных показателей по реальным временным рядам (в частности, показателей Ляпунова, корреляционной размерности – этим проблемам посвящено большое количество работ), используемых в теории детерминированного хаоса, при определении которых используется предельный переход. При работе с конечным рядом обеспечить предельный переход невозможно. Поэтому во многих случаях переходят к сравнительным исследованиям показателей, полученных с помощью стандартных измерительных процедур.

На основе теоретических подходов с учетом условий и ограничений работы с реальными временными рядами разработаны и представлены в литературе способы и модели обработки нелинейных сигналов, включающие в себя методы классификации и сравнения временных рядов, выявления свойств динамических систем, лежащих в их основе, предсказания будущих значений – экстраполяции временных рядов. Краткое описание основных принципов некоторых из них так же изложено в первой главе.

Описание и обзор литературы по двум основным подходам к анализу и прогнозу нерегулярных временных рядов на основе глобальных и локальных моделей вынесено во ***вторую главу***. Деление методов на глобальные и локальные проводится по области определения параметров аппроксимирующей функции. В глобальных методах эти параметры идентифицируются с использованием всех известных значений ряда. Локальная аппроксимация предполагает отказ от явного использования для прогнозирования всех уже известных значений ряда и ограничивает количество объясняющих значений лишь наиболее близкими в некотором смысле к стартовой точке, после которой начинается прогноз.

Основное направление использования глобальных методов – это получение глобальных характеристик системы. Прогнозирование в этих методах используется в большой степени для выяснения долгосрочной динамики, чем для оценки ближайшего будущего. Локальные методы прогноза имеют преимущество в задачах, связанных с прогнозированием нерегулярных временных рядов.

«Глобальный» метод – сингулярный спектральный анализ (SSA) и метод локальной аппроксимации (LA) детально рассмотрены во второй главе. Там же даны подробные пошаговые описания этих методов.

В основе большинства подходов, связанных с обработкой временных рядов $\{x_1,\ldots,x_N\}$, лежит построение множества векторов задержек $\mathbf{x_t}=\begin{pmatrix} x_t & x_{t-1} & \cdots & x_{t-p+1}\end{pmatrix}^{\mathrm{T}}$, $t=p,p+1,\ldots,N$. Метод задержек устанавливает переход от исходного одномерного (скалярного) временного ряда к многомерному (векторному) представлению. При этом каждый многомерный вектор образуется из некоторого числа p следующих друг за другом значений исходного ряда. Результат можно представить в виде набора «фотографий» ряда, сделанных через скользящее вдоль ряда окно, в которое одновременно попадает лишь p последовательных значений ряда:

$$\begin{array}{c} \\ \\ \\ \mathbf{X}_{p\times(N\text{-}p+1)}= \\ \\ \\ \end{array}\begin{array}{c} \vdots \\ x_{p+1} \\ \left(\begin{bmatrix} x_p \\ \vdots \\ x_2 \\ x_1 \end{bmatrix}\right. \\ \downarrow \\ \\ \end{array}\begin{array}{c} \vdots \\ x_{p+2} \\ \begin{bmatrix} x_{p+1} \\ \vdots \\ x_3 \\ x_2 \end{bmatrix} \\ x_1 \\ \downarrow \end{array}\begin{array}{c} \\ \\ \cdots \\ \\ \\ \end{array}\begin{array}{c} \\ \\ \left.\begin{bmatrix} x_N \\ \vdots \\ x_{N-p+2} \\ x_{N-p+1} \end{bmatrix}\right). \\ x_{N-p} \\ \vdots \end{array}$$

Принцип действия SSA во многом схож с Фурье-фильтрацией: здесь исходный ряд также представляется в виде набора составляющих, только в SSA эти составляющие не являются в общем случае гармоническими. Особенностью метода SSA является обработка матрицы $\mathbf{X}$ по алгоритму, близкому к методу главных компонент (МГК). Использование для обработки всей матрицы $\mathbf{X}$ сразу обуславливает отнесение этого метода к разряду глобальных.

Выбор составляющих осуществляется из условия максимизации разброса точек (каждый столбец матрицы $\mathbf{X}$ представляется в виде точки в M-мерном пространстве задержек) вдоль выбранной составляющей. Иллюстрация этого алгоритма представлена на рис. 1, где $x^{(1)},x^{(2)},x^{(3)}$ – первая, вторая и третья оси координат исходного базиса, $\left(y^{(1)},y^{(2)},y^{(3)}\right)$ – новый базис. Ось $y^{(1)}$ расположена вдоль прямой максимального разброса точек на графике; по оси $y^{(2)}$ разброс меньше, но он максимальный из всех возможных при заданной

$y^{(1)}$; ось $y^{(3)}$ при заданных $\left(y^{(1)}, y^{(2)}\right)$ определятся единственным образом и на ее долю разброса почти не остается.

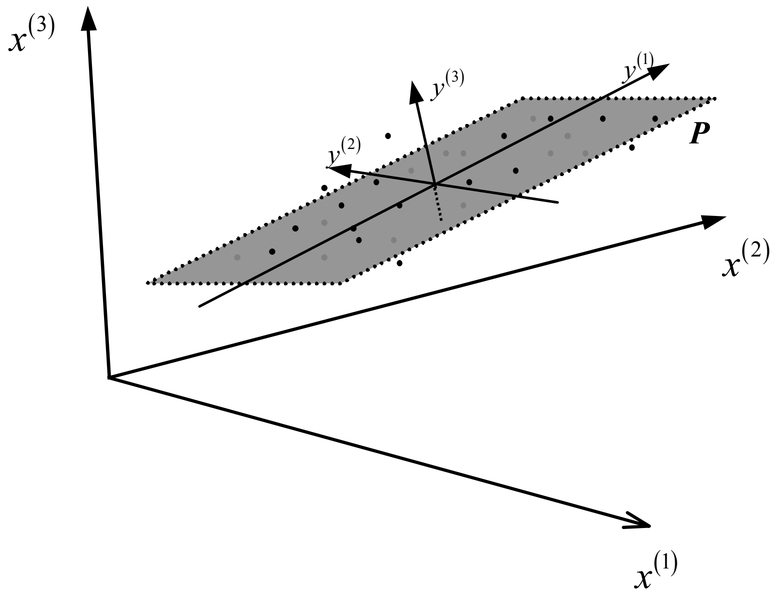

Рис. 1. Выбор новой системы координат методом главных компонент ($M = 3$).

В SSA получающееся разложение используется для выделения наиболее значимых составляющих ряда и отсева случайных возмущений. Например, для случая, изображенного на рис. 1, можно ограничиться первыми двумя составляющими, так как вдоль третьей оси разброса почти нет и все точки можно приблизительно считать лежащими в плоскости двух первых составляющих – (P). Отклонения от этой плоскости как раз могут быть следствием случайных возмущений.

Анализ полученных составляющих позволяет выделять периодические и квазипериодические составляющие временного ряда. Этот метод может использоваться для улучшения соотношения сигнал/шум. Кроме того, в последнее время появились оригинальные варианты, расширяющие возможности SSA и позволяющее строить на его основе прогноз дальнейшей динамики ряда. Например, метод экстраполяции «Гусеница», в соответствии с которым предсказание ряда на один шаг по времени вперед определяется из условия минимизации проекции нового вектора на выбранную гиперплоскость (P).

Иллюстрацией алгоритма *локальной аппроксимации* в простейшем одномерном случае может быть следующий способ построения прогноза температуры на следующий день: сначала находится день, в который температура была максимально близка к сегодняшней и затем в качестве прогноза температуры на завтра берется ее значение в день, следующий за найденным. Аппроксимация более высокого порядка позволяет также учитывать влияние на прогноз отклонений «сегодняшней» температуры от температуры в день, наиболее похожий на сегодняшний.

Построение прогноза на один шаг по времени по методу LA проводится в три этапа. Сначала строится матрица задержек и выбирается локальное представление, т. е. вид функции, связывающей следующее значения ряда с предыдущими: $x_{t+1} = f\left(\mathbf{x_t}, \mathbf{a}\right)$, где $\mathbf{a}$ – вектор параметров представления.

Наиболее распространенный вариант – линейная аппроксимация (аппроксимация первого порядка), но используются еще два варианта: нулевого (аппроксимация константой) и второго порядков (аппроксимация полиномом второго порядка).

Затем производится выбор соседей – векторов, ближайших к последнему известному вектору в пространстве задержек (но не во времени). После этого производится оценка параметров представления исходя из известной динамики векторов-соседей. Оценив значения параметров аппроксимации, можно построить прогноз следующего значения ряда: $\hat{x}_{L+1} = f\left(\mathbf{x}_L, \hat{\mathbf{a}}\right)$. Индексом L обозначен последним известный (стартовый) вектор.

Для прогноза на несколько шагов обычно используются один из двух способов: итеративный и прямой. В работе предложен еще один способ – итеративный с пересчетом.

Итеративный способ состоит в последовательном построении прогноза на один шаг с добавлением его результата к исходным данным и повторным применением модели представления с параметрами, оцененными на первом шаге. Идея итеративного способа с пересчетом состоит в том, чтобы заново оценивать параметры представления на каждом шаге. Как показано в работе, это позволяет существенно повысить эффективность метода.

При прямом способе прогноза стартовый вектор и все его соседи остаются неизменными, а параметры представления оцениваются заново на каждом шаге: $\hat{x}_{L+t} = f\left(\mathbf{x}_L, \hat{\mathbf{a}}^{(t)}\right)$. Здесь не требуется заново выбирать соседей и не происходит накопления ошибки за счет итераций.

Во второй главе на тестовых примерах дано предварительное качественное сравнение результатов получаемых при разных порядках аппроксимации при прогнозировании на один шаг вперед и разных способах аппроксимации при прогнозировании на большую длину.

Построению общей математической модели метода локальной аппроксимации и выбору оптимального варианта метода для условий конкретного временного ряда посвящена ***третья глава***.

Предлагаемая в этой главе модель позволяет получить аналитической вид решения задачи прогноза методом LA. Исследование свойств этого решения позволяет выработать рекомендации по выбору способа прогноза (итеративный, итеративный с пересчетом, прямой) без проведения численных расчетов. Кроме того, на основе этого решения становится возможным изучение асимптотических свойств прогноза и критериев выбора порядка локальной аппроксимации.

Как показано в работе, уравнение LA любого порядка (локальное представление) может быть представлено в виде: $x_{t+1} = a_0 + \mathbf{x}_t{}^{\mathrm{T}} \mathbf{a}$, где $\mathbf{a}$ – обобщенный вектор коэффициентов – параметров модели.

Для оценки этих коэффициентов строится система уравнений $\mathbf{Y} = \mathbf{I} a_0 + \mathbf{X}\mathbf{a}$, где $\mathbf{I}_{\Xi\times 1}$ – вектор из единиц, $\mathbf{X}$ – матрица соседей – транспонированный фрагмент матрицы задержек, состоящий лишь из одних соседей стартового вектора, а $\mathbf{Y}$ – вектор значений в которые переходят соседи стартового вектора за Υ шагов. Требуется оценить параметры $\left(a_0, \mathbf{a}\right)$.

Решение этой системы уравнений было получено с помощью метода наименьших квадратов (МНК). В результате была определена формула для прогнозируемого значения ряда: $\hat{x}_{L+\Upsilon} = \overline{Y} + \left(\mathbf{x}_L^{\mathrm{T}} - \overline{\mathbf{X}}\right)\hat{\mathbf{a}}$, где $\overline{Y}$ — среднее значение вектора $\mathbf{Y}$ (вектор, совпадающий с результатом прогноза для нулевого порядка), $\overline{\mathbf{X}}$ — вектор средних значений координат в матрице соседей $\mathbf{X}$, т. е. «усредненный» сосед.

Как установлено в работе, прогноз, полученный методом локальной аппроксимации любого порядка, есть линейная комбинация прогноза нулевого порядка и отклонений стартового вектора от «усредненного» соседа. Отсюда можно сделать вывод, что поправка, обусловленная отклонением стартового вектора от «усредненного» соседа, корректирует неравномерность распределения соседей вокруг стартового вектора.

На основании общего аналитического решения были получены модельные уравнения прогноза для каждого из рассматривавшихся способов прогноза (итеративного, итеративного с пересчетом, прямого).

Было установлено, что итеративный вариант нулевого порядка дает аппроксимацию нулевого порядка, т. е. прогноз на любое количество шагов аппроксимируется значением прогноза на один шаг вперед. В то же время при прямом способе построения прогноза нулевого порядка сразу получается аппроксимация первого порядка, а при равномерном распределении соседей вокруг стартового вектора – второго порядка. Таким образом, при использовании прямого способа нулевого порядка прогноз тем точнее, чем ближе и равномернее относительно стартового вектора распределены соседи, тогда как итеративный способ подобен стоящим часам, которые иногда показывают точное время.

Для выбора оптимального способа прогнозирования на несколько шагов вперед были исследованы асимптотические свойства прогноза при итеративном и прямом способах прогноза. Установлено, что при $\Upsilon \to \infty$ прямой способ дает более приемлемые результаты, чем итеративный.

С той же целью была выполнена оценка ошибки прогноза, связанной с конечной точностью вычислений. Оказалось, что при итеративном способе экстраполяции ошибка не меньше, а, как правило, больше, чем при прямом.

Основным результатом проведенного сравнения стало установление факта, что наиболее универсальным методом при прогнозировании достаточно длинных стационарных нерегулярных временных рядов является прямой вариант LA нулевого порядка. С одной стороны, он может обеспечивать прогнозирование с величиной ошибки, убывающей пропорционально среднему квадрату отклонений соседей от стартового вектора. С другой стороны, расчет по этому методу является самым простым из всех рассмотренных. В случае ограниченного набора данных или их неравномерного распределения вокруг стартового вектора можно воспользоваться прямым вариантом LA первого порядка.

Представленные в третьей главе результаты численного сравнения подтвердили основные выводы, сделанные при анализе аналитического решения: из рассмотренных примеров следует, что для рядов, ограниченных по количеству исходных данных, в каждом случае лучшим оказался первый порядок аппроксимации.

При использовании различных способов прогноза худшие результаты получаются при итеративном способе. Итеративный способ с пересчетом дает в среднем результаты, сопоставимые по точности с прямым вариантом, однако реально он может применяться только в нулевом порядке, поскольку при использовании его старших порядков за счет погрешностей численных расчетов возможно появление неконтролируемых ошибок прогноза, значительно превосходящих среднее значение ряда.

В ***четвертой главе*** представлены результаты, полученные при использовании методов сингулярного спектрального анализа и локальной аппроксимации в обработке реальных временных рядов. Там также предложен способ подавления шума в обрабатываемом ряде и критерий автоматического выбора параметров локальной аппроксимации.

В качестве объекта исследований использовалась реальная последовательность чисел Вольфа, характеризующих солнечную активность. Интерес к исследованию солнечной активности связан, кроме прочего, с заметной корреляцией пиков активности Солнца с проявлениями общественной активности. Достаточно отметить, что последние три максимума активности Солнца пришлись на 1979-1980, 1989 и 2000 годы. При этом даже весьма подробные модели не позволяют точно спрогнозировать момент и величину очередного пика активности Солнца.

Несмотря на сравнительно небольшую длину последовательности чисел Вольфа метод SSA позволил выявить ее компоненты, отвечающие уже известным солнечным циклам. В то же время SSA не позволяет точно оценить моменты наступления пиков активности (появления максимальных чисел Вольфа), хотя и дает при этом весьма точную оценку их величины. В целом можно сказать, что описанный метод SSA представляется достаточно эффективным и весьма перспективным методом предсказания динамики магнитной активности Солнца.

В случае сильно зашумленных временных рядов (к ним относятся почти все ряды внелабораторного происхождения) даже увеличение числа наблюдений не позволяет эффективно применять алгоритм LA, так как в этом случае высока вероятность появления большого количества ложных соседей и отсев соседей истинных.

Для целей обработки зашумленных временных рядов в работе предложено соединить оба рассмотренных метода (SSA и LA) в один, в котором SSA использовался бы только для фильтрации исходного временного ряда (подавления шума), а сам прогноз строится по методу LA.

Получившийся метод (SSA-LA) можно рассматривать как расширение метода LA, позволяющее применять его к сильно зашумленным временным рядам. В рамках этого метода для построения прогноза с учетом предшествующего анализа предложен прямой вариант LA первого порядка и SSA с центрированием.

Иллюстрация результатов применения этого метода при разных уровнях шума представлена на рис. 2. Исходя из этих результатов, можно ожидать, что предварительная SSA-фильтрация позволяет значительно повысить точность и устойчивость прогноза, получаемого методом LA. Причем такой результат не зависит от уровня шума, длины прогноза и системы, породившей исследуемый временной ряд.

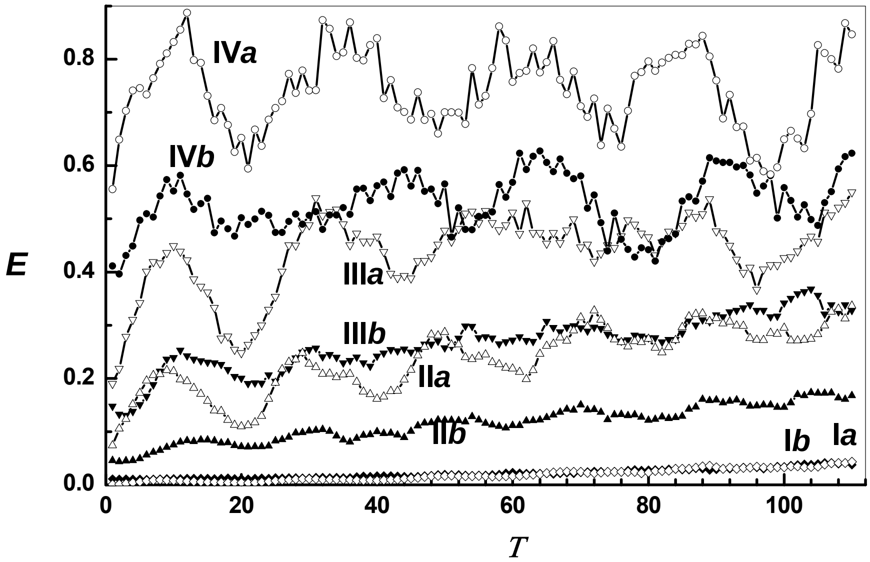

Рис. 2. Ошибка прогноза методами LA (a) и SSA-LA (b) для уравнения Маккея-Гласса в зависимости от интервала прогнозирования в отсутствие шума (I) и при амплитудах шума 1.5% (II), 5% (III), 20% (IV). Медианное усреднение по 500 стартовым точкам.

Основная техническая проблема, возникающая при прогнозировании реальных временных рядов методом LA (равно как и SSA) связана с выбором размерности вложения p и количества соседей. Для оптимизации выбора этих параметров в работе предложен критерий минимизации ошибок автопрогнозов соседей:

$$\sigma_{\text{u}}^2 = \frac{\sum_{n=1}^{\Xi}\left(\hat{x}_{s_n+\Upsilon} - x_{s_n+\Upsilon}\right)^2}{\Xi - p} \to \min_{p,\Xi}.$$

Здесь для каждого значения размерности вложения из заданного интервала и каждого числа соседей из своего интервала оцениваются коэффициенты $\hat{\mathbf{a}}$, задающие переход на Υ шагов вперед. Затем по этим коэффициентам рассчитываются «будущие» значения соседей: $\hat{x}_{s_n+\Upsilon} = \hat{x}_{L+\Upsilon} + \left(\mathbf{x}_{s_n}^{\mathrm{T}} - \mathbf{x}_{L}^{\mathrm{T}}\right)\hat{\mathbf{a}}$. Окончательно выбирается тот набор соседей и размерность вложения, при которых достигается минимум величины σ_{u}^2.

Применение предложенного критерия позволят определять параметры аппроксимации без эмпирического исследования различных вариантов прогноза, поэтому такой способ определения параметров может быть легко реализован программно и использоваться без необходимости визуального контроля, т.е. практически в автоматическом режиме.

Использование рассматриваемого критерия имеет еще два полезных приложения: возможность определения максимальной длины прогноза и возможность оценки числа соседей. При этом анализ найденных соседей может использоваться для выяснения характерного периода (если он есть) рассматриваемого временного ряда. Для этого достаточно проанализировать расстояния во времени между соседями.

Возможности алгоритма автоматического выбора параметров в методе локальной аппроксимации (автоLA) проиллюстрированы на тестовом и реальном примерах.

Результаты прогноза солнечной активности методами автоLA и SSA представлены на рис. 3. Черными кружками показаны реальные значения чисел Вольфа до середины 2004 года. Прогноз построен по данным до 1989 года. Остальные значения использовались для проверки качества прогноза. Светлой линией на графике показан прогноз методом автоLA. Критерий определения максимальной длины прогноза показал, что дальше 1998 года прогноз автоLA не обоснован. Поэтому дальнейшие результаты, представленные лишь для сравнения с SSA, показаны штриховой линией. Сравнивая прогнозы LA и SSA можно отметить, что SSA точнее выделяет периодические составляющие и дает более точный долгосрочный прогноз для достаточно регулярных рядов (к которым относится и ряд чисел Вольфа), тогда как LA более эффективен для краткосрочных прогнозов.

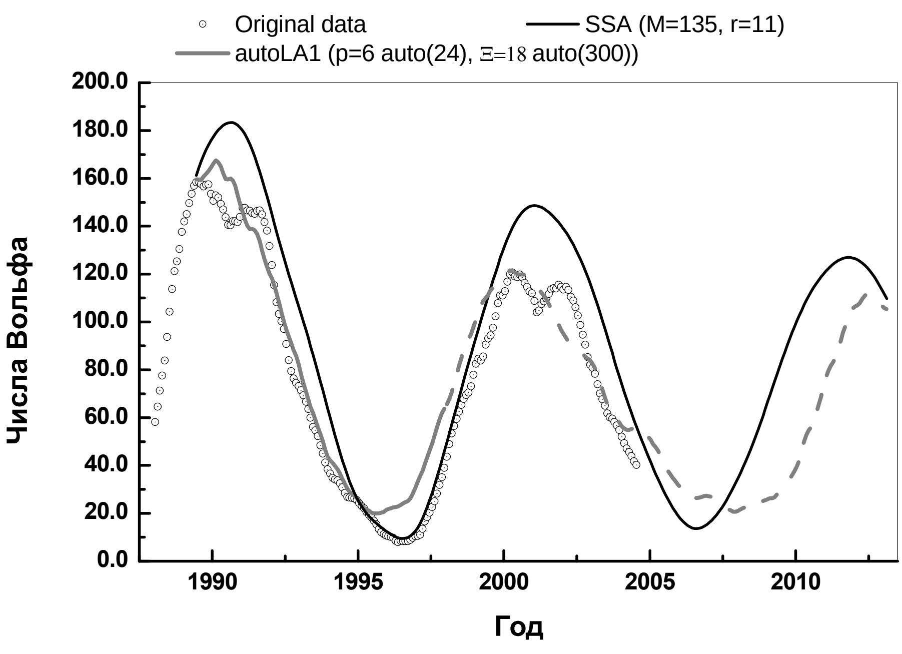

Рис. 3. Прогноз солнечной активности.

Заключение

Методы и подходы, выдвинутые в рамках теории динамических систем, уже дано вышли за пределы самой дисциплины. Одним из важных направлений практического приложения теории динамических систем стали разработанные в ее рамках модели анализа временных рядов.

Основные результаты работы, которые выносятся на защиту.

1. Аналитически и на реальных данных изучены возможности и ограничения методов сингулярного спектрального анализа и локальной аппроксимации.
2. Разработана общая математическая модель метода локальной аппроксимации, на основе которой удалось выяснить существенные особенности получаемого решения задачи прогноза и предложить способ выбора варианта метода с учетом условий конкретной задачи и объема имеющихся данных.
3. Предложен способ предварительной фильтрации зашумленных временных рядов, позволяющий существенно повысить надежность прогноза.
4. Обоснован алгоритм автоматического выбора параметров локальной аппроксимации, который обеспечивает возможность использования метода локальной аппроксимации в составе программного комплекса. Алгоритм позволяет отказаться от визуального контроля при выборе параметров метода.
5. На основе построенных моделей и алгоритмов разработан программный комплекс для анализа и прогнозирования временных рядов, реализующий, в том числе полностью автоматизированный режим обработки с контролем обоснованности прогноза.

Приведенные результаты позволяют снять основное ограничение, сдерживающее распространение метода локальной аппроксимации в исследованиях реальных временных рядов, связанное с необходимостью фактически интуитивного выбора параметров и варианта метода для условий решаемой задачи. Это в свою очередь может способствовать значительному расширению сферы применения метода локальной аппроксимации.

Автор выражает искреннюю и глубокую признательность коллегам и друзьям за помощь при подготовке диссертации.

Публикации

1. *Лоскутов А.Ю., Истомин И.А., Котляров О.Л., Кузанян К.М.* Исследование закономерностей магнитной активности Солнца методом сингулярного спектрального анализа // Письма в Астрономический журнал. — 2001. №11. — Т. 27. — С.867-876.
2. *Loskutov A., Istomin I.A., Kuzanyan K.M. and Kotlyarov O.L.* Testing and forecasting the time series of the solar activity by singular spectrum analysisи // Nonlin. Phenomena in Complex Syst. — 2001. — Vol. 4, №1. — P.47-57.
3. *Лоскутов А.Ю., Котляров О.Л., Истомин И.А., Журавлев Д.И.* Проблемы и методы нелинейной динамики. III. Локальные методы прогнозирования временных рядов // Вестн. Моск. ун-та. Физ.-Астрон. — 2002, № 2. — С.3-21.
4. *Лоскутов А.Ю., Котляров О.Л.* Нелинейная динамика и анализ временных рядов // Проблемы анализа риска. — 2004, № 2. — Т. 1. — С.160-177.
5. *Истомин И.А., Котляров О.Л., Лоскутов А.Ю.* К проблеме обработки временных рядов: расширение возможностей метода локальной аппроксимации посредством сингулярного спектрального анализа // Теоретическая и математическая физика — 2005, № 1. — Т. 142 — С.148-159.
6. *Лоскутов А.Ю., Котляров О.Л., Журавлев Д.И.* Временные ряды: анализ и прогноз // Сб. научных трудов 11-й Международной конф. «Математика, компьютер, образование» — Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», 2004. —Т. 1. — С.9-46.
7. *Loskutov A., Istomin I.A., Kuzanyan K.M. and Kotlyarov O.L.* Testing and forecasting time-series of the Solar activity by singular spectrum analysis // http://xxx.lanl.gov/ps/nlin/0010027.
8. *Loskutov A., Istomin I. and Kotlyarov O.* Data analysis: generalizations of the local approximation method by singular spectrum analysis // http://xxx.lanl.gov/abs/nlin.cd/0109022.